AI and Problem Solving: Emotions

Jocelyn Paine

June 3, 1997

This tutorial is concerned mainly with theories of emotion and their necessity for successful action in the real world. As such, it follows from the tutorial on Planning. Most of the reading will have to be done in the RSL.

1 Introduction — emotions and planning

I'll start with two popular articles. First, that by Keith Oatley in *New Scientist* for 19th August 1989. This introduces the main idea: how emotion acts as a global "redirecting" device. Following that, chapter 20 of *The Computer and The Mind* by Philip Johnson-Laird (1989), which expands on the same theme.

2 Followup — emotions and planning

Start with *Why robots will have emotions*, in *International Joint Conference on AI 7* (1981). There's a photocopy of this in Psychology.

Most of the rest of the reading is from the journal *Cognition and Emotion* (RSL Psychology, open shelves). Start with the article by Johnson-Laird in number 1, volume 1 of the journal. It is, I think, the first time these ideas were published in their present form, though Sloman's earlier article makes some of these points, and Simon's book *The Science of the Artificial* contributed some ideas.

Johnson-Laird's article is followed by a criticism from Aaron Sloman. Read that as well. (You may find that these two are unavailable because the journal is being bound. If so, sorry. I've put copies of some of the other references in Psychology.)

The first article refers to the Hacker planning program, and the difference between "careful" and "careless" modes. Look these up in *AI and Natural Man.*

You will find the same topics further developed in volume 1 number 3 of *Cognition and Emotion.* Read the editorial and Sloman's article; and at least skim Frijda's article. There are photocopies of Sloman's and Frijda's articles in Psychology.

Do you think Frijda's program is in any sense a cognitive model? Does it actually have emotions? Or it it just a superficial simulation of behaviour?

3 Consciousness and self-monitoring

Most of these articles have referred to the idea that consciousness arises from the need to monitor one's mental tasks, allocating appropriate resources to each. This is related to the idea that consciousness arises whenever an organism creates a mental model of its own operation. Obviously both are relevant to the topic of limited resources in planning; and they would be of great philosophical interest if they had something compelling to say about how consciousness arose. I don't believe they have; but read them anyway, because I may be wrong and it's an important topic.

The easiest-to-find work on this is the final chapter of *Mental Models* by Johnson-Laird (1983). See also *A computational analysis of consciousness* by him in *Cognition and Brain Theory* volume 6 page 4499. There's a photocopy of that in Psychology.

You should also look at the references on Miles Glen's and Peter McLeod's reading lists for Consciousness (in the H.I.P. reading lists file, by the catalogue in the Psychology library).

Why am I unconvinced about self-monitoring entailing consciousness? For one reason, because I know a number of programs that do, and I don't believe they're conscious! (I'm using one at the moment as I edit). The best-known class of such programs is *operating systems*, the programs that allocate time between the many users using a big computer like the VAX. These were probably the source of the idea of self-monitoring in the works above. To see what such programs do, and by analogy, what it's thought the mind does, read the article on *Time-sharing systems* by Fano and Corbato (especially pages 133–135) in *Scientific American* for September 1966. You may also find the article by Samuel in *New Scientist* for May 27th 1965 (volume 26, no 445) helpful.

4 Emotions and story-understanding

In the Planning tutorial, I asked you to read a paper suggesting that we need planners to understand other people's plans as well as to create our own. That was chapter 26 of *Thinking — Readings in Cognitive Science*, edited by P.N.Johnson-Laird and P.C.Wason. There was a more recent account in Chapter 10 of *Artificial Intelligence* by Charniak and McDermott, concerning attribution of emotions to the characters.

Perhaps an understander needs emotions too. The article on *Emotional modelling* in the *Encyclopaedia of AI* gives a brief account of this. *Beyond the Schema Given* in *Cognition and Emotion* number 1 volume 3 is the most accessible reference.

5 Are these models biologically plausible?

Most of the papers above look at emotion as a necessary switching mechanism that would have to be properly integrated with a general-purpose planner, diverting its focus when necessary. It's a nice idea, and may help us in building better programs. But — considering that the physiological linkages on which emotion signals are based came long before rational thought — is it biologically plausible?

Consider what I asked you to read in the *Unconventional Approaches* section of the planning tutorial. The approach there is to build up through the evolutionary tree, making artificial organisms that mimic the competences of real animals. These animals do not represent goals, subgoals, tasks, plans, symbolically in the same way that conventional planners do. Consequently, it would be impossible to build emotional switching on to them. Yet, if they're based with sufficient accuracy on real animals, they'll embody some mechanism corresponding to emotion. Where does this leave Oatley, Johnson-Laird and Sloman? No references yet, but think about it.

6 Other miscellaneous work on emotions and cognition

I don't know of any other work to which we have access on the topics above. For completeness, read the rest of the *Encyclopaedia of AI* article. If you're interested, read up on Parry and the artificial neurotic in *AI and Natural Man.* I shan't set Finals questions on them (though someone else might), but they're interesting as early cognitive simulations, designed to test specific theories.

7 Essay

Will an intelligent robot be emotional? Will it be aware of its emotions?